

How real can they get? Investigating neural responses to GAN generated scenes.

Aylin Kallmayer & Melissa L.-H. Vo

Scene Grammar Lab, Department of Psychology, Goethe University Frankfurt, Germany

To understand how we efficiently navigate real-world scenes, we need to unravel the underlying computations and structure of representations that afford efficient scene processing. One hypothesis is that we exploit scene structures by learning hierarchical object-to-object and scene-to-object relations captured by a *scene grammar*. But how can these high-level networks be learnt? Does unsupervised learning automatically lead to representations that reflect properties of scene grammar? To assess how well scenes generated by generative adversarial networks (GANs) capture real-world scene structure perceived over time we conducted an EEG experiment. Participants viewed 180 generated scenes across six categories (30 exemplars/category). Generated scenes varied in their “realness” as assessed by three different measures: realism ratings, false-alarm (FA) rates, and categorization performance for 50 and 500ms presentation times. While ratings and FAs served as explicit and implicit measures of a scene’s general realism, respectively, categorization performance was a more direct measure of how well generated scenes capture scene category specific information. Using multivariate pattern analysis (MVPA) we were able to decode scene category from neural responses to generated images with peak performances around 140 and 640ms. This suggests that generated scenes evoke scene category specific information during early and late processing. To test whether we could predict our behavioral measures with neural responses over time, we ran ridge regularized regressions for each timepoint. Realism ratings as well as FAs in the 50ms condition were best predicted by neural signals around 330ms. Surprisingly, we could not predict categorization performance for generated scenes from the neural signal. From this we conclude that information contained in generated scenes that makes them appear “real” is neurally processed around 330ms, while actual categorization performance of generated scenes could not be predicted by these neural signatures implying a lack of category specific scene structure usually captured by scene grammar.

299/300

Acknowledgements: This work was supported by SFB/TRR 26 135 project C7 to Melissa L.-H. Vö and the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project ‘The

Adaptive Mind') and the Main-Campus-Doctus stipend awarded by the Stiftung Polytechnische Gesellschaft to Aylin Kallmayer.